

The use of DNA polymorphisms in population history and adaptive variation studies in pines

(Wykorzystanie polimorfizmu DNA w badaniach historii populacji i zmienności adaptacyjnej sosen)

Wachowiak Witold
Cavers Stephen
Żukowska Weronika
Wójkiewicz Błażej



Przegląd

***P. mugo complex* i *P. sylvestris* jako eksperymentalny system w badaniach lokalnej adaptacji**

Analiza polimorfizmu nukleotydowego
(struktury populacji, identyfikacji loci poddanych działaniu selekcji)

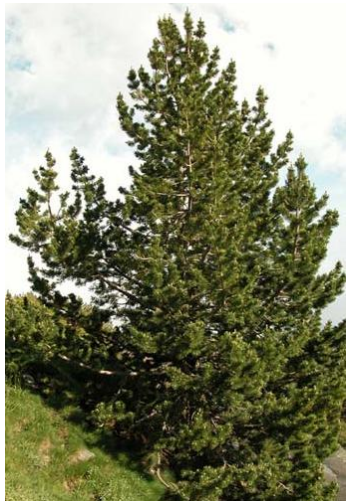
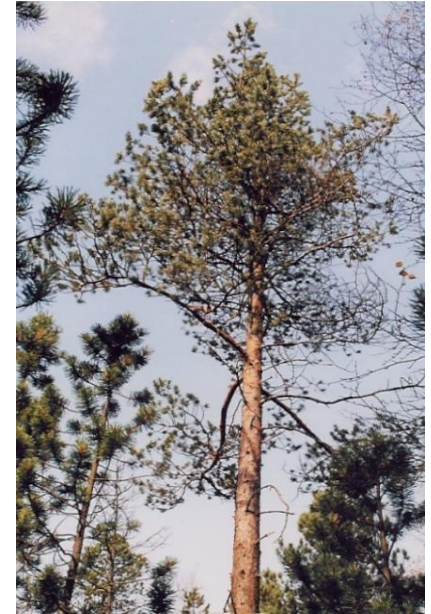
Nowe zasoby genomowe

Pinus mugo complex i *P. sylvestris*

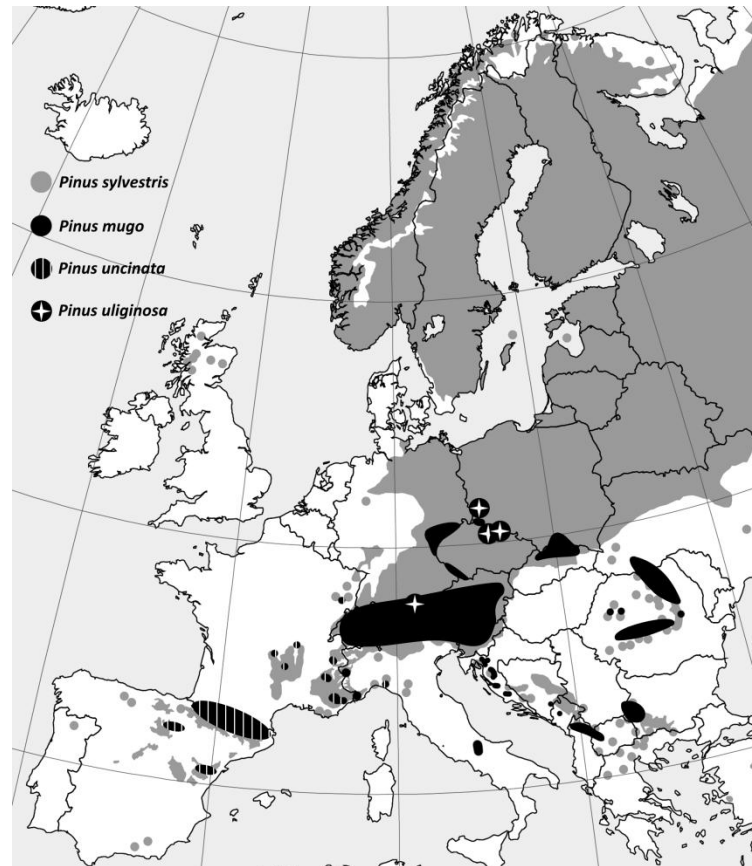
P. mugo



P. sylvestris



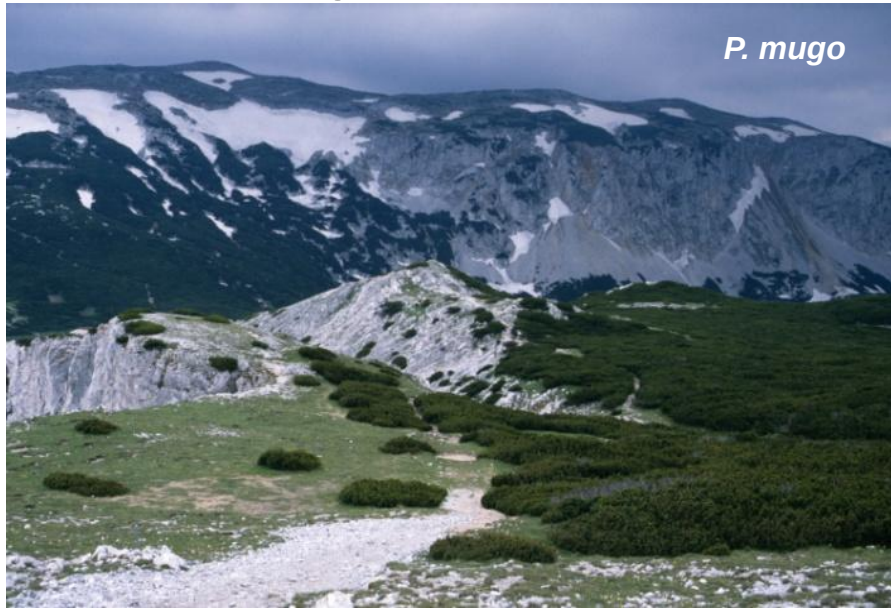
P. uncinata

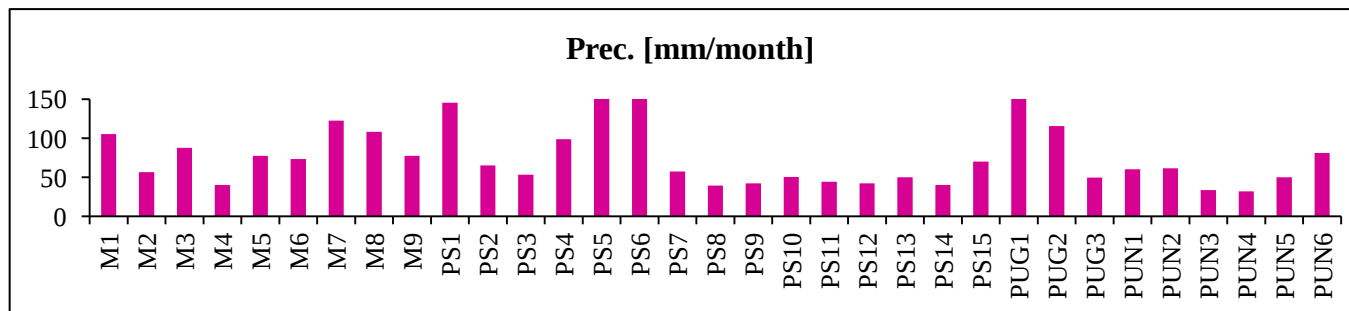
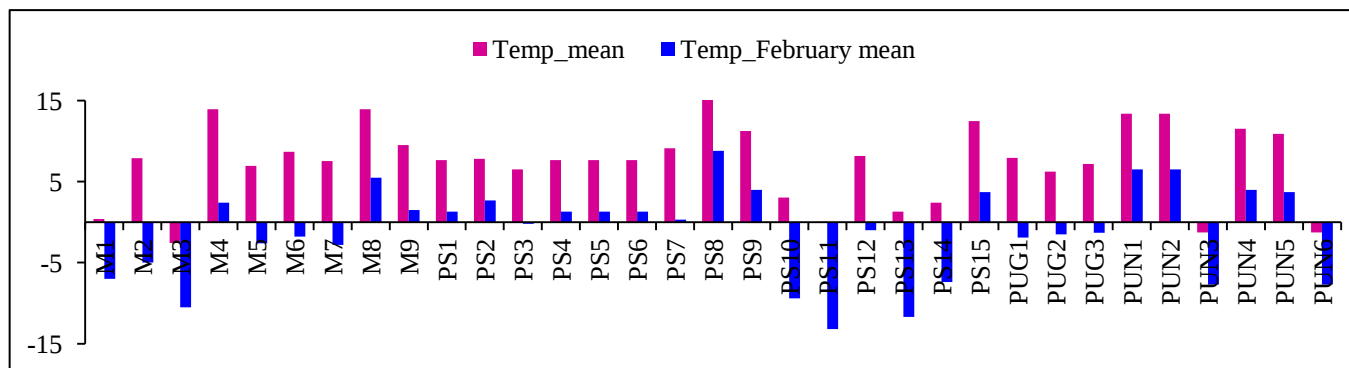
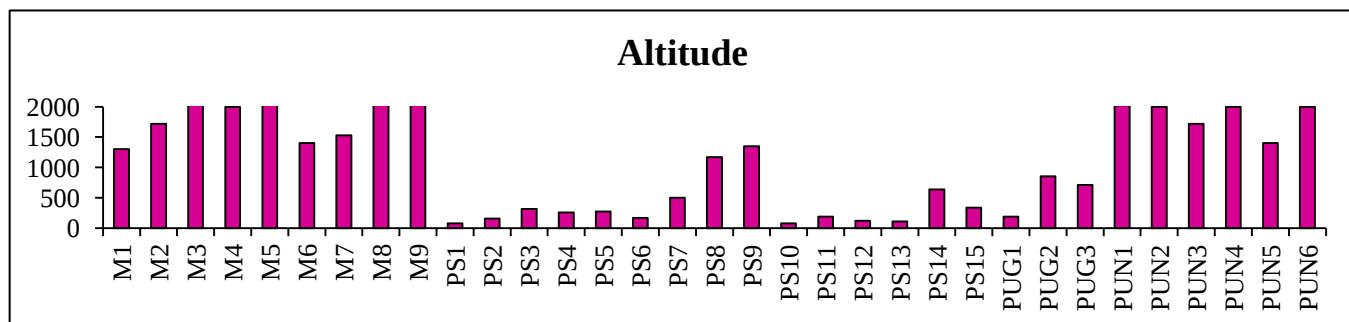
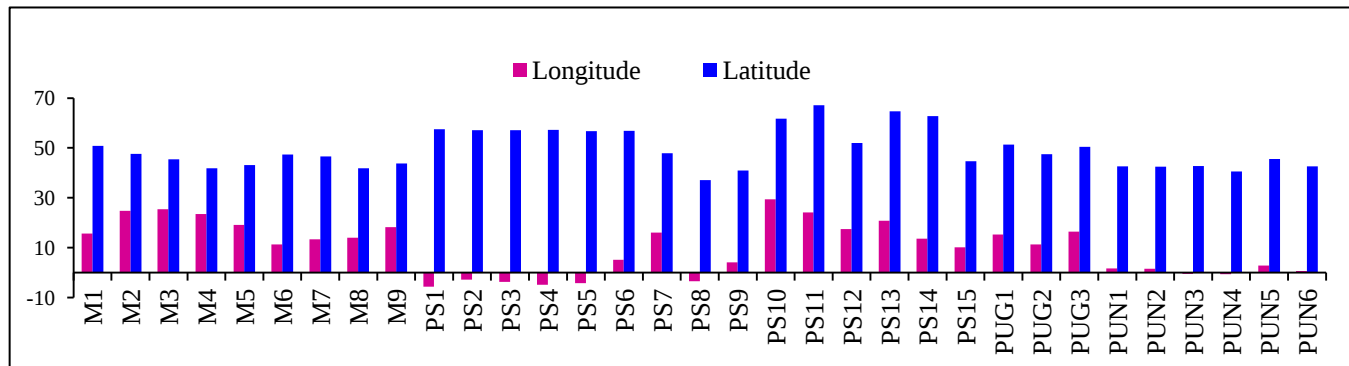
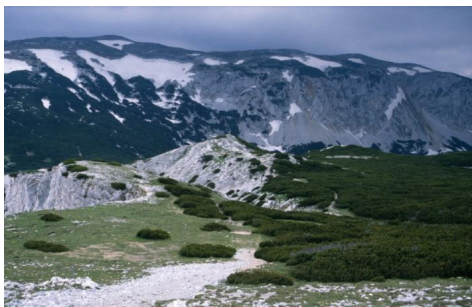


P. uliginosa



Gradient czynników środowiskowych w zasięgu występowania





Pytanie natury ogólnej:

Czy potrafimy zidentyfikować geny poddane działaniu naturalnej selekcji potencjalnie związane z różnicowaniem się tych taksonów i lokalną adaptacją ?

Cechy genomu warunkujące podejście badawcze (przynajmniej dotychczas...)

- genom ogromnych rozmiarów (~ 25 000 MBP)
- niski poziom LD

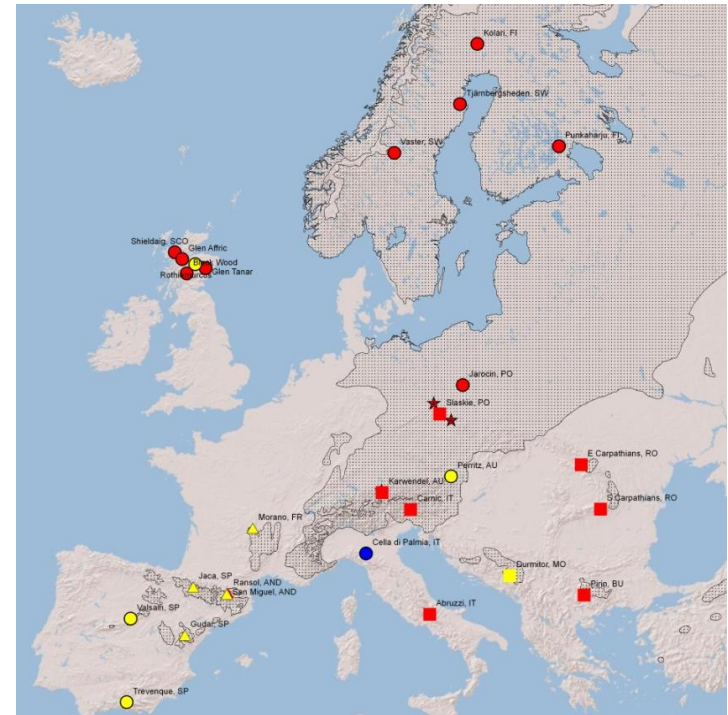
⇒ ograniczenie na poziomie analiz całego genomu

⇒ analiza ściśle określonych rejonów (np. CG) i próba korelacji polimorfizmu ze zmiennością np. fenotypową, geograficzną, ekologiczną ...



Analiza struktury populacji i zmienności adaptacyjnej 4 gatunków sosen

- analiza polimorfizmu ~100 genów nDNA
(genes related to metabolisms, transport, expression regulation, signal transduction etc.)
- 30 populacji z zasięgu Europejskiego

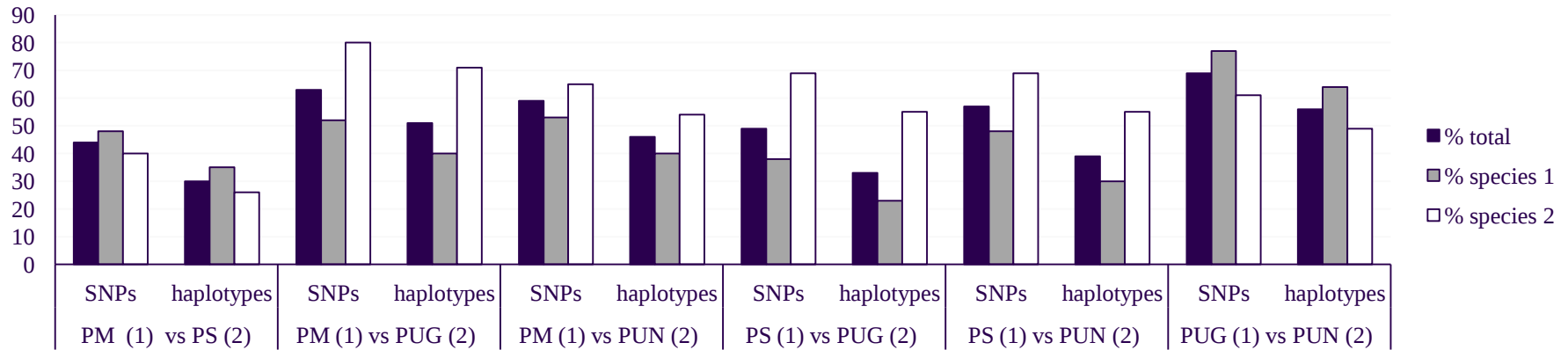


Pytania:

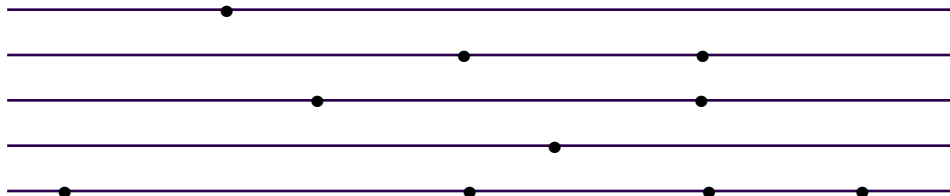
- ⇒ Jaki jest poziom zróżnicowania genetycznego pomiędzy gatunkami?
- ⇒ Jaka jest struktura populacji ?
- ⇒ Czy potrafimy zidentyfikować geny, których zmienność jest wynikiem działania naturalnej selekcji.

Poziom polimorfizmu nukleotydowego

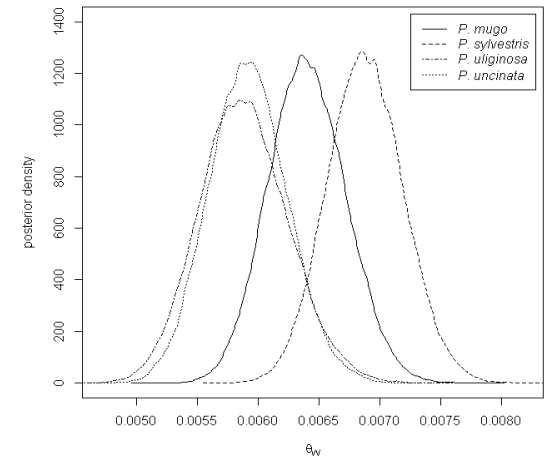
- homogenne tło genetyczne, niska dywergencja (<1%)



- jednakowy polimorfizm nukleotydowy



Multilocus $\theta_w = 4Ne\mu = 0.006-0.007$
(Watterson 1975)



Brak dowodów na istnienie struktury populacji

Analiza na podstawie 1680 SNPs

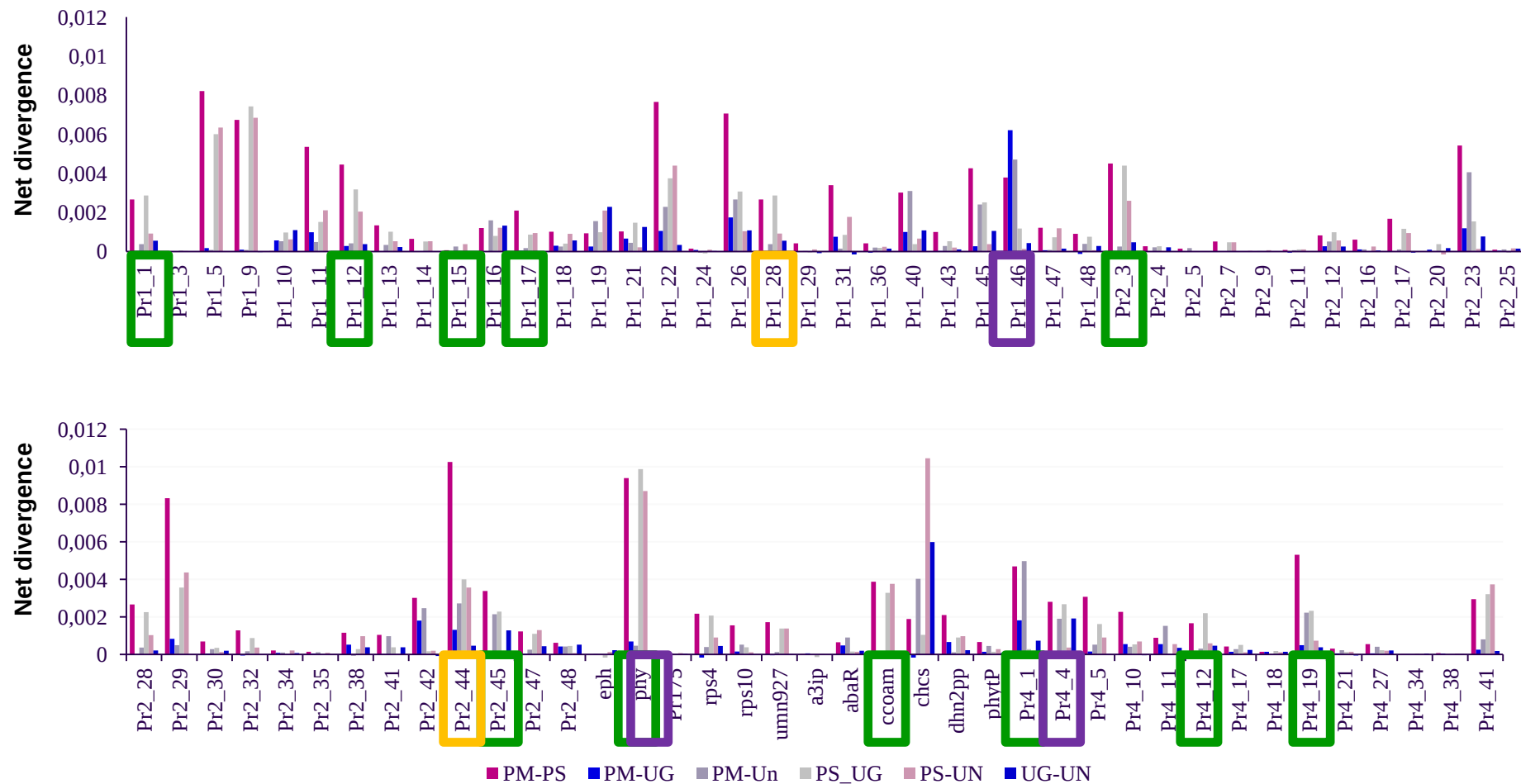


P. mugo (80 individuals)
main mountain chains in Europe

P. sylvestris (135 individuals)
south – north of Europe

P. uliginosa (24) *P. uncinata* (50)
EU_C Iberian Peninsula

Geny poddane działaniu selekcji - potencjalnie w wyniku lokalnej adaptacji populacji



⇒ **polimorfizm genów zgodny z modelem selekcji w wyniku lokalnej adaptacji**

Accron.	Gene function/name [category]	<i>P. mugo</i>	<i>P. sylvestris</i>	<i>P. uliginosa</i>	<i>P. uncinata</i>
Pr1_1	putative beta-alanine ligase [M]		X		
Pr1_12	NAD(P)-linked oxidoreductase protein [M]		X		
Pr1_15	mitogen-activated protein-kinase [ST]		X		
Pr1_17	glutamate transporter [T]		X		
Pr1_28	translation initiation factor-4G-like [E]			X	
Pr1_46	alpha-N-acetylglucosaminidase [M]	X			
Pr2_3	2-3 ethylene-resp.transcription factor 1B [E]		X		
Pr2_44	DNA repair helicase XPB1-like [M]			X	
Pr2_45	hypothetical protein [UN]		X		
ccoam	Caffeoyl CoA O-methyltransferase [M]		X		
eph	putative epoxide hydrolase [M]		X		
phy	putative phytocyanin [T]	X	X		
Pr4-1	Peroxidase [ST]		X		
Pr4-5	calcium dependent proteokinase [ST]	X			
Pr4-12	proton myo-inositol transporter [T]		X		
Pr4-19	laccase [M]		X		

Wachowiak W., Telford A., Zucca GM, González-Martínez SC, Cavers.

Signatures of local adaptation and divergence in four closely related pine species.

Heredity, submitted

Dostępne zasoby genomowe do analiz polimorfizmu i struktury genetycznej populacji

~ 100 genów nDNA - 1680 SNPs (projekt NERC)

Loci mikrosatelitarne cpDNA i nDNA (35 loci) (projekt NCN)

Vendramin i in.(1996), Provan i in.(1999), Sebastiani i in.(2012), Elsik i Williams (2001), Soranzo i in. (1998), Elsik et al. (2000), Zhou i in. (2002)

<u>cpDNA 1</u>	<u>cpDNA 2</u>	<u>nDNA 1</u>	<u>nDNA 2</u>	<u>nDNA 3</u>	<u>nDNA 4</u>
Pt15169	PCP1289	psyl44	PtTX3116	ptTX 3107	PtTX4001
Pt26081	PCP26106	psyl42	Spac 11.6	Spag 7.14	PtTX4011
Pt30204	PCP30277	psyl2	Spac 12.5	PtTX2146	Spac 11.4
Pt36480	PCP36567	psyl57	Spac 11.8		PtTX3032
Pt71936	PCP41131	psyl16	psyl17		PtTX3025
	PCP45071	psyl25			
	PCP87314	psyl19			
	PCP102652	psyl18			
		psyl36			

Wójkiewicz B, Żukowska W, Wachowiak W. Multiplexed chloroplast and nuclear mikrosatellite loci for genetic studies in pines. W przygotowaniu

Opracowanie nowych zasobów genomowych – sekwencjonowanie transkryptomu (Illumina)

Ekstrakcja RNA dla 17 osobników

(*P. sylvestris* 5, *P. mugo* 5, *P. uncinata* 4, *P. uliginosa* 3)

Spectrum™ Plant Total RNA Kit (Sigma)

biblioteki cDNA

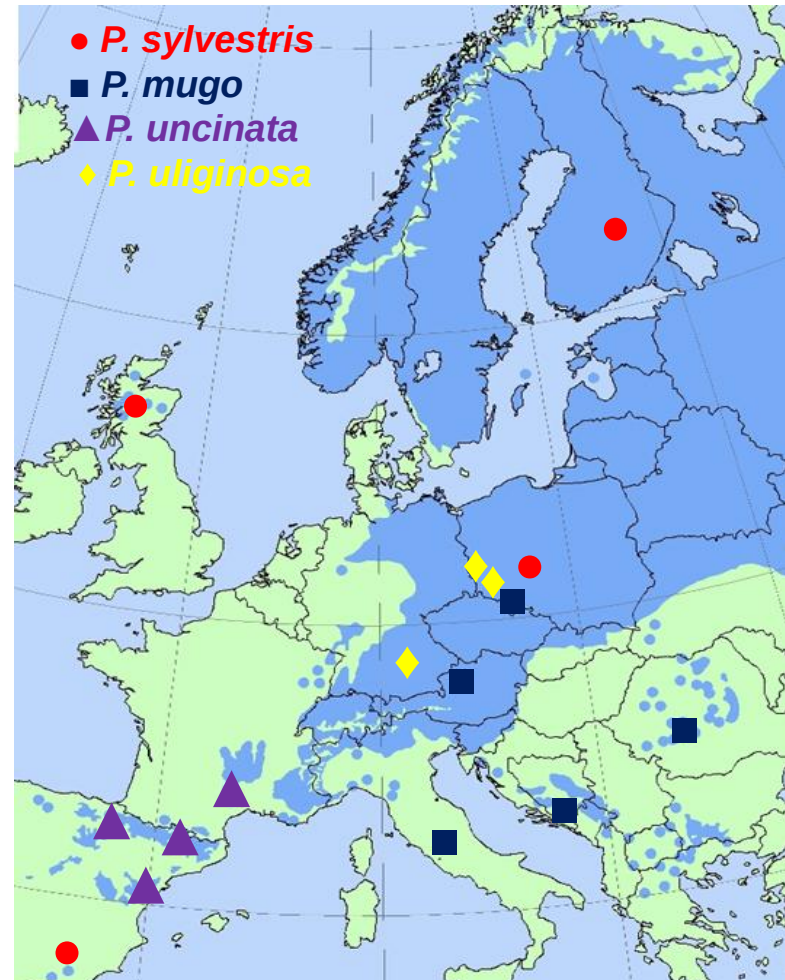
(® *TruSeq™ RNA Sample Preparation Kit (Illumina)*)

100bp pair-end Sequencing

(*Illumina HiSeq 2000 platform at Edinburgh Genomics, the University of Edinburgh, Scotland*)

Złożenie *de novo* transkryptomu dla

P. sylvestris (*Trinity - version r2012-06-08*)

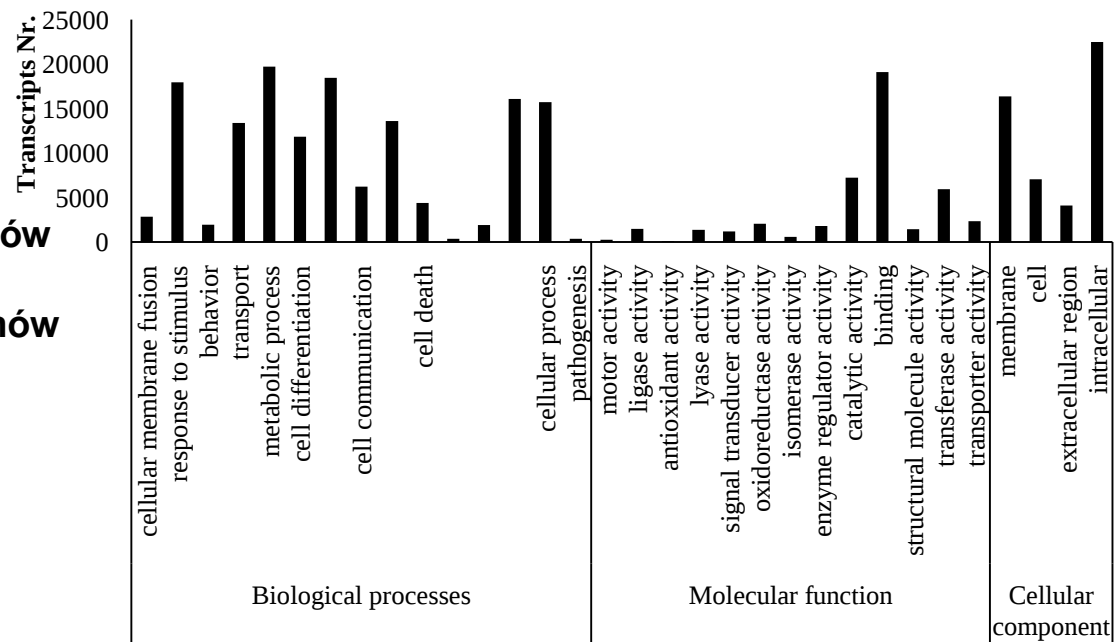


Tranksryptom *P. sylvestris*

Assembly metric	Raw assembly	Unigene set
Num contigs >100	151932	40968
Total bases in contigs >100	119,849,194	61,246,267
GC contigs >100	41.8	42.5
Mean length for contigs >100	788.8	1501.2

Podstawowe statystyki :

- 40968 transkryptów
- 40798 z wyłączeniem retrotranspozonów
- przypisane do 19659 unikatowych genów i znanych białek
- 9529 genów przypisanych do 2130 ścieżek enzymatycznych



Statystyki mapowania:

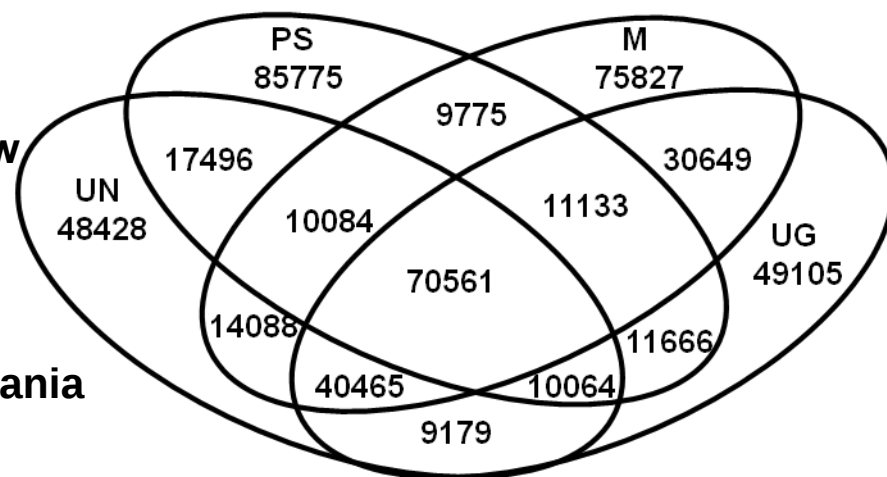
Sample Nr,	Species	Total reads	% Mapped reads	Number of SNPs
1.	<i>P. sylvestris</i>	31,116,472	84.14%	67817
2.	<i>P. sylvestris</i>	258,401,512	85.34%	81519
3.	<i>P. sylvestris</i>	37,849,782	80.21%	63874
4.	<i>P. sylvestris</i>	38,970,706	79.47%	94021
5.	<i>P. sylvestris</i>	45,140,044	81.73%	95814
6.	<i>P. mugo</i>	43,752,078	78.50%	116762
7.	<i>P. mugo</i>	32,600,000	78.40%	103818
8.	<i>P. mugo</i>	40,104,880	78.19%	100602
9.	<i>P. mugo</i>	52,934,684	78.67%	130942
10.	<i>P. mugo</i>	69,248,828	80.03%	138989
11.	<i>P. uncinata</i>	34,805,254	79.34%	102037
12.	<i>P. uncinata</i>	30,291,214	78.06%	87181
13.	<i>P. uncinata</i>	45,834,740	77.85%	118800
14.	<i>P. uncinata</i>	48,550,644	79.12%	115050
15.	<i>P. uliginosa</i>	52,718,596	78.55%	127068
16.	<i>P. uliginosa</i>	40,729,720	80.38%	116297
17.	<i>P. uliginosa</i>	69,457,322	79.37%	147646

Statystyki SNPs :

• SNPs zidentyfikowane w 54-59% trankryptów

(u 22041 genów dla *P. sylvestris*, 24096 dla *P. mugo*,
22416 dla *P. uncinata* oraz 22710 dla *P. uliginosa*)

• 259,087 SNPs odpowiednich do genotypowania (odległość pomiędzy SNPs minimum 50bp)

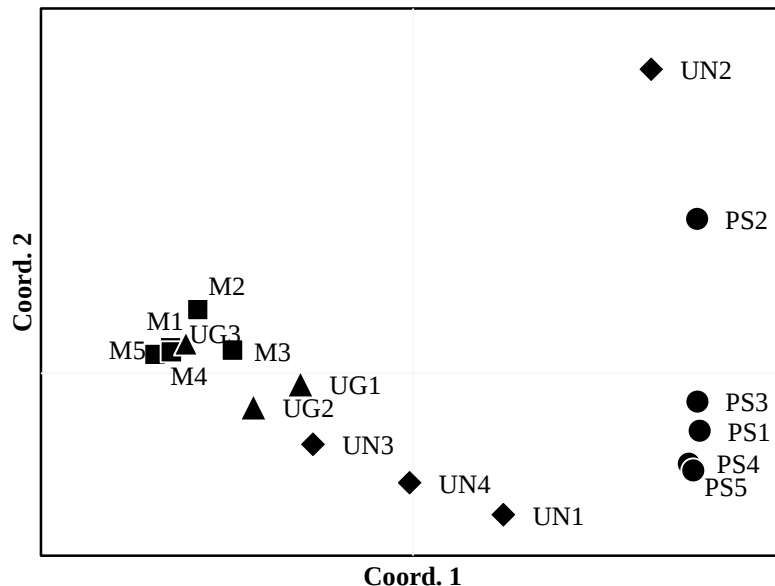


Polimorfizm nukleotydowy dla 676 homologicznych rejonów genomowych (*n*DNA)

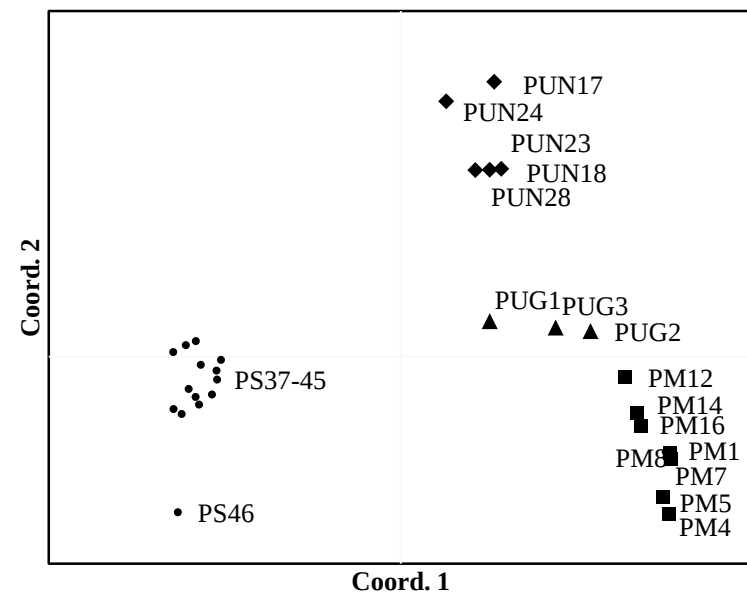
Species	N	L	SNPs	Sing.	Π_{tot}	D
<i>P. sylvestris</i>	5	1364676	12920	8674	0.0044	-0.243
<i>P. mugo</i>	5	1364676	13129	8710	0.0045	-0.221
<i>P. uncinata</i>	4	1364676	13420	10374	0.0053	-0.169
<i>P. uliginosa</i>	3	1364676	9581	9581	0.0047	-
Total/Aver.	17	1364676	27929	12181	0.0047	-0.211

Analiza PCA:

676 contigów - 27,929 SNPs



79 genów kandydatów - 1680 SNPs



Biorą pod uwagę fakt:

• dane z analiz sekwencjonowania DNA nowej generacji nie dostarczają zwyczajnie większej liczby miejsc polimorficznych poprzez dostęp do większych fragmentów genomu

• wielkość i charakterystyka genomu szpakowych

→ większość SNPs przydatna do genotypowania

Research article

Highly accessed

Open Access

Comparative transcriptomics of a complex of four European pine species

Witold Wachowiak^{1,2*}, Urmi Trivedi³, Annika Perry¹ and Stephen Cavers¹

* Corresponding author: Witold Wachowiak witcho@ceh.ac.uk

► [Author Affiliations](#)

For all author emails, please [log on](#).

BMC Genomics 2015, **16**:234 doi:10.1186/s12864-015-1401-z

Published: 25 March 2015

Abstract (provisional)

Background *Pinus sylvestris*, *P. mugo*, *P. uliginosa* and *P. uncinata* are closely related but phenotypically and ecologically very distinct European pine species providing an excellent study system for analysis of the genetic basis of adaptive variation and speciation. For comparative genomic analysis of the species, transcriptome sequence was generated for 17 samples collected across the European distribution range using Illumina paired-end sequencing technology. Results De novo transcriptome assembly of a reference sample of *P. sylvestris* contained 40968 unigenes, of which fewer than 0.5% were identified as putative retrotransposon sequences. Based on gene annotation approaches, 19659 contigs were identified and assigned to unique genes covering a broad range of gene ontology categories. About 80% of the reads from each sample were successfully mapped to the reference transcriptome of *P. sylvestris*. Single nucleotide polymorphisms were identified in 22041-24096 of the unigenes providing a set of ~220-262 k SNPs identified for each species. Very similar levels of nucleotide polymorphism were observed across species ($n=0.0044-0.0053$) and highest pairwise nucleotide divergence (0.006) was found between *P. mugo* and *P. sylvestris* at a common set of unigenes. **Conclusions** The study provides whole transcriptome sequence and a large set of SNPs to advance population and association genetic studies in pines. Our study demonstrates that transcriptome sequencing can be a very useful approach for development of novel genomic resources in species with large and complex genomes.

The complete article is available as a **provisional PDF**. The fully formatted PDF and HTML versions are in production.

Podziękowania



NATIONAL SCIENCE CENTRE
POLAND

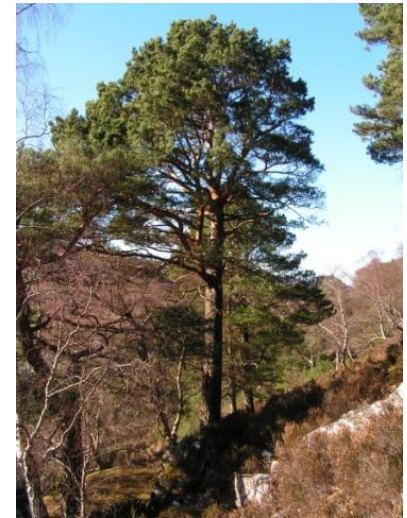
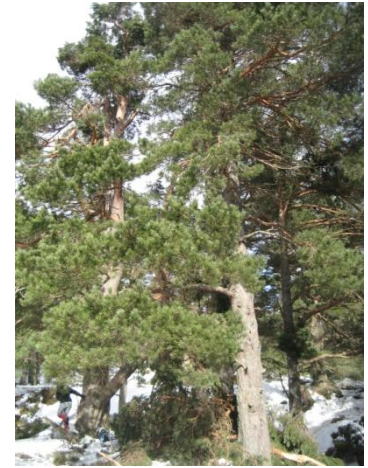


Ministry of Science
and Higher Education

Republic of Poland



NATURAL
ENVIRONMENT
RESEARCH COUNCIL



Thank you!